

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平6-318094

(43)公開日 平成 6年(1994)11月15日

(51)Int.Cl.⁵

G 1 0 L 3/00
5/04

識別記号

庁内整理番号

H 8946-5H
F 8946-5H

F I

技術表示箇所

審査請求 未請求 請求項の数 1 O L (全 6 頁)

(21)出願番号 特願平5-106683

(22)出願日 平成 5年(1993) 5月 7日

(71)出願人 000005049

シャープ株式会社

大阪府大阪市阿倍野区長池町22番22号

(72)発明者 木村 治

大阪府大阪市阿倍野区長池町22番22号 シ
ャープ株式会社内

(72)発明者 海木 延佳

大阪府大阪市阿倍野区長池町22番22号 シ
ャープ株式会社内

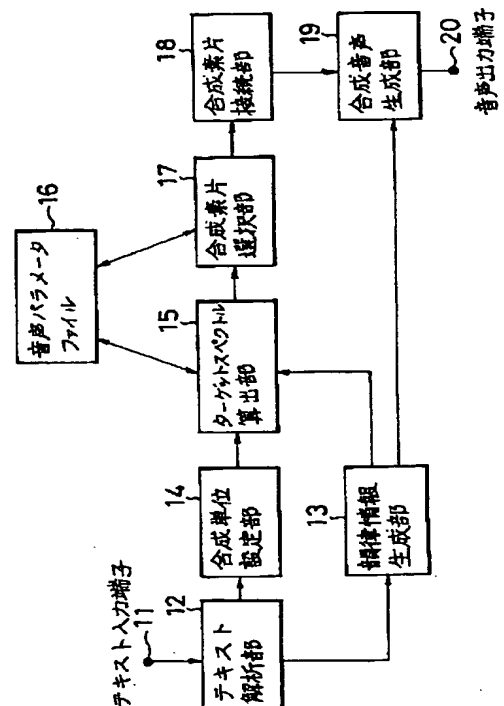
(74)代理人 弁理士 川口 義雄 (外 1 名)

(54)【発明の名称】 音声規則合成装置

(57)【要約】

【目的】 大量の音声データから演算量が比較的少なく、しかも素片接続部のスペクトル歪みの少ない素片を選択することにより、明瞭性及び自然性が高い合成音声を出力できる音声規則合成装置を提供することにある。

【構成】 自然音声を分析して音韻毎にラベル付けされた音声合成パラメータを格納する音声パラメータファイル (16) と、出力音声を組み立てるために適切な合成単位の情報を設定する合成単位設定部 (14) と、音韻情報、音律情報、及び合成単位設定部で設定された情報に基づいて合成単位での接続部におけるターゲットスペクトルを算出するターゲットスペクトル算出部 (15) と、ターゲットスペクトル算出部で算出されたターゲットスペクトルに基づいて音声パラメータファイルに格納されている音声合成パラメータから適切な合成素片を選択する合成素片選択部 (17) と、合成素片選択部で選択された合成素片を接続する合成素片接続部 (18) とを備えている。



【特許請求の範囲】

【請求項1】 自然音声进行分析して音韻毎にラベル付けされた音声合成パラメータを格納する記憶手段と、出力音声を組み立てるために適切な合成単位の情報を設定する設定手段と、音韻情報、音律情報、及び該設定手段で設定された情報に基づいて該合成単位での接続部におけるターゲットスペクトルを算出する算出手段と、該算出手段で算出された該ターゲットスペクトルに基づいて該記憶手段に格納されている該音声合成パラメータから適切な合成素片を選択する選択手段と、該選択手段で選択された該合成素片を接続する接続手段とを備えていることを特徴とする音声規則合成装置。

【発明の詳細な説明】

【0001】

【産業上の利用分野】 本発明は、合成音声を生成する音声規則合成装置に関する。

【0002】

【従来の技術】 規則に従って音声合成する従来の音声合成装置では、音声の合成単位として音韻や、音節、VCV（母音・子音・母音）接続、CVC（子音・母音・子音）接続、単語など音韻との対応や、調音結合を考慮した単位を設定し、自然音声进行分析して作成した音声合成パラメータ値を記憶しておき、入力文字列に対応する単位の音声合成パラメータ（以下、合成素片と呼ぶ）の編集、結合、変形により音声合成していた。

【0003】

【発明が解決しようとする課題】 しかしながら、上述した従来の音声合成装置では、同じ音素や音節で、単位毎に発声して集めた音と文章中に現れる音がかなり異なるため、合成音の自然さに欠けるという問題点があった。

【0004】 例えば、単音節などを発声した自然音声进行分析したもので文章の音声合成すると、一音一音ははっきりと発音しているような印象の合成音になってしまう。合成音の速度をあげるほどその傾向が強い。

【0005】 また、あらかじめ文章や単語のように合成単位よりも長い単位で発声した自然音声を大量に持ち、最適な素片を選択して合成素片として用いると、調音結合はすでに表現されているので自然性が向上するが、最適な素片を選択する規則がまだ見いだされていない。

【0006】 特に、合成素片の接続による歪みを少なくするために、合成素片の接続部のスペクトル歪みを考慮して素片を選択するには、素片間のスペクトル間距離を算出する必要があり、素片の組合せの多さから多大の演算量が必要であるという問題点があった。

【0007】 本発明の目的は、大量の音声データから演算量が比較的少なく、しかも素片接続部のスペクトル歪みの少ない素片を選択することにより、明瞭性及び自然性が高い合成音声出力できる音声規則合成装置を提供することにある。

【0008】

【課題を解決するための手段】 本発明の目的は、自然音声进行分析して音韻毎にラベル付けされた音声合成パラメータを格納する記憶手段と、出力音声を組み立てるために適切な合成単位の情報を設定する設定手段と、音韻情報、音律情報、及び設定手段で設定された情報に基づいて合成単位での接続部におけるターゲットスペクトルを算出する算出手段と、算出手段で算出されたターゲットスペクトルに基づいて記憶手段に格納されている音声合成パラメータから適切な合成素片を選択する選択手段と、選択手段で選択された合成素片を接続する接続手段とを備えている音声規則合成装置によって達成される。

【0009】

【作用】 本発明の音声規則合成装置では、設定手段は、音節やVCV（母音・子音・母音）音韻系列など出力音声を組み立てる上で適切な合成単位を設定し、算出手段は、音韻情報と韻律情報および上記合成単位設定部からの情報により上記合成単位での接続部におけるターゲットスペクトルを算出し、記憶手段は、大量の自然音声进行分析して音韻毎にラベル付けされた音声合成パラメータ値を格納し、選択手段は、算出手段からの情報により、記憶手段より適切な合成素片を選択し、接続手段は、選択された合成素片を接続する。

【0010】

【実施例】 以下、図面を参照して、本発明の音声規則合成装置の実施例を説明する。

【0011】 図1は、本発明の音声規則合成装置の一実施例の構成を示すブロック図である。

【0012】 図1の音声規則合成装置は、テキスト入力端子11に接続されておりテキスト入力端子から入力された変換すべきテキストを基に形態素解析、漢字かな変換、アクセント処理等を行なって出力するテキスト解析部12、テキスト解析部12に接続されておりテキスト解析部12から出力された解析情報を基にピッチパターン、各音素毎の時間長パターン、及び振幅パターンを生成して出力する韻律情報生成部13、テキスト解析部12に接続されておりテキスト解析部12から出力された解析情報を基に出力音声を組み立てるために合成単位に分割して出力する設定手段である合成単位設定部14、韻律情報生成部13及び合成単位設定部14に接続されており合成単位設定部14で合成単位に分割された前後の音韻系列と韻律情報生成部13からの情報を基に最適なターゲットスペクトルを算出して出力する算出手段であるターゲットスペクトル算出部15、ターゲットスペクトル算出部15に接続されており大量の音声データを基に合成に必要な音響パラメータを分析、作成するして出力する記憶手段である音声パラメータファイル16、ターゲットスペクトル算出部15及び音声パラメータファイル16に接続されており合成素片の接続部のスペクトルがターゲットスペクトルに最も近いものを音声パラメータファイル16の中から選択して出力する選択手段であ

る合成素片選択部17、合成素片選択部17に接続されており選択された素片同士を結合して出力する接続手段である合成素片接続部8、韻律情報生成部13及び合成素片接続部18に接続されており合成素片接続部18で得られた合成素片系列及び韻律情報生成部13で得られた韻律情報を基に合成音声を生成して出力端子20に出力する合成音声生成部19によって構成されている。

【0013】次に、図1の音声規則合成装置の動作を説明する。

【0014】テキスト入力端子11より音声に変換すべきテキストが入力されると、テキスト解析部12より係り受けなどの構文解析や品詞解析などの形態素解析、及び漢字かな変換、アクセント処理が行われ、合成単位設定部14、韻律情報生成部13に必要な解析情報が送出される。その解析情報としては合成単位設定部14に対しては音韻の区別を示す記号列、韻律情報生成部13に対しては呼気段落内モーラ数、アクセント形、発声スピードなどである。

【0015】韻律情報生成部13は、これらの情報を基にピッチパターン、各音素毎の時間長パターン、及び振幅パターンを規則により生成する。

【0016】合成単位設定部14は、入力された音韻記号列を、音節やVCV音韻系列など出力音声を組み立てる上で適切な合成単位に分割し、その分割された音韻系列をターゲットスペクトル算出部15に出力する。

【0017】ターゲットスペクトル算出部15は、合成単位に分割された前後の音韻系列と、韻律情報生成部13からの情報を基に最適なターゲットスペクトルを算出する。

【0018】音声パラメータファイル16は、大量の音声データを基にオフライン処理であらかじめ作成しておく。例えば、アナウンサー人による単語、文章など数時間分の音声データに対しデジタルソナグラムによる視察により音韻ラベリングを施して、合成に必要な音響パラメータを分析しておく。

【0019】合成素片選択部17は、合成素片の接続部のスペクトルが、上記ターゲットスペクトルに最も近いものを音声パラメータファイル16の中から選択する。

【0020】合成素片接続部18は、選択された素片どうしの結合を行なって合成波形生成部19に送出する。

【0021】合成音声生成部19は、合成素片接続部18で得られた合成素片系列と、韻律情報生成部13で得られた韻律情報を基にして合成音声を生成し、生成した音声出力端子10に出力される。

【0022】上述した構成では、テキスト解析部12を設けているが、あらかじめテキスト解析を行い、その解析情報を本装置へ入力した場合には、テキスト解析部12を省略できる。

【0023】同様に、あらかじめ韻律のパターンを生成し本装置へ入力した場合は、韻律情報生成部13を省略で

きる。

【0024】ここで用いる音響パラメータ及び合成音声を生成するための合成器については、特に規定するものではなく全てに対して適用可能である。

【0025】次に、図2のフローチャートを参照して、上記ターゲットスペクトル算出部15の動作を詳細に述べる。

【0026】図2は、／oNsei／を合成する場合の／N／のターゲットを算出する一例を示している。

【0027】まず、前後の合成単位と韻律情報を入力し（ステップS1）、接続部の音韻を中心に音韻系列を設定し（ステップS2）、音声パラメータファイル16からその音韻系列を含む音声パラメータを検索する（ステップS3）。

【0028】もし、候補が見つからない場合は、順次検索音韻系列を両側から削除しながら検索を行なう。例えば／oNse／を含む音声パラメータがないときは、／oNs／→／Ns／→／N／となる。

【0029】次に、韻律情報から接続部のピッチ条件を設定し（ステップS4）、候補の絞り込みを行なう（ステップS5）。このピッチ条件は、例えばピッチの±5%などとする。もし該当するものがなければ、ピッチ条件を±10%、15%……と広げていく。

【0030】次に、候補の中から接続部の音韻の継続長に最も近いものを選択し（ステップS6）、選択された音声パラメータから接続音韻の中心のスペクトルを算出し（ステップS7）、ターゲットスペクトルとする（ステップS8）。

【0031】以上の処理で接続部のターゲットスペクトルを算出する。

【0032】次に、図1の音声規則合成装置による音声規則の合成処理を具体的に説明する。

【0033】例えば「音声」という単語がテキスト入力端子11に入力されると、テキスト解析部12で／oNsei／という音韻系列と韻律情報が生成される。そして、合成単位をVCVとすると、合成単位設定部14で／So／、／oN／、／Nse／、／ei／、／iS／の5つの合成単位に分割される。ただし、／S／は無音をあらわす。次に合成単位毎に素片を選択して行くが、以下に／Nse／の場合の例を示す。

【0034】まず、ターゲットスペクトル算出部15で／oN／と／Nse／の接続部のターゲットスペクトルを算出する。この場合、／oNsei／の音韻系列の音声パラメータを音声パラメータファイル16から検索し、韻律情報からの絞り込みによって選択された音声パラメータの／N／の時間的中心であるスペクトルをターゲットスペクトルSP1とする。

【0035】同様に／Nse／と／ei／との接続部のターゲットスペクトルSP2も算出する。

【0036】次に、合成素片選択部17で、／Nse／

5

の音韻系列を持つ音声パラメータを音声パラメータファイル16から検索する。次に、ターゲットスペクトルSP1、SP2と検索された候補毎に／N／及び／e／の部分のスペクトル距離の最小値を算出し、その最小値の和が最も小さい候補を合成素片として選択する。このようにして合成単位毎に素片を選択した後、合成素片の接続をターゲットスペクトルとの距離が最小の位置で行ない、合成波形を生成する。

【0037】このように接続部における最適なターゲットスペクトルを設定し、これに最も近いスペクトルを持つ合成素片を接続していくことによって、接続歪みの少ない合成音声を得られる。

【0038】従来のターゲットスペクトルを設定しないで接続歪みの少ない合成を行なう方式では、接続する合成素片間の組合せの多さのために多大の計算量を要していたのに対し、本装置では計算量の大幅な削減が可能である。

【0039】更に、計算量及びメモリを削減する方法として、音韻系列及び韻律情報毎にあらかじめターゲットスペクトルを算出し、そのターゲットスペクトルに最適な合成素片をテーブル登録しておく。

【0040】例えば、VCV単位の合成でハツオン／N／も母音として考えると、接続部は／a、i、u、e、o、N／の6種類である。

【0041】最小のハード構成を考えると、あらかじめ普通の高さで発声した単母音の定常部を分析しておき、それぞれのスペクトルをターゲットスペクトルとする。

【0042】次に、上記合成素片選択部17と同様のアルゴリズムで、ターゲットスペクトルに最適な合成素片を選択し、これをテーブル登録しておく。そして合成時には、そのテーブルを参照することによって合成素片を選択する。この場合、VCV毎に1種類の合成素片が対応しているテーブルを構築できる。

【0043】この方法では、合成時に検索処理を行なう方法に比べて合成音の品質が落ちる可能性はあるが、テーブルに記述された合成素片のみを音声パラメータファイルにメモリするだけでよく、更に合成時に検索処理を行なわないので、計算量及びメモリを大幅に削減できる。

6

【0044】また、もう少し大きなハード構成が可能なら、複数の高さで発声した単母音の定常部をターゲットにしたり、調音結合の影響を強く受ける音韻系列（例えば無声化や鼻音化）の音声からターゲットを作成し、合成素片テーブルを作成することによって更に高品質化をはかることができる。

【0045】

【発明の効果】本発明の音声規則合成装置は、自然音声を分析して音韻毎にラベル付けされた音声合成パラメータを格納する記憶手段と、出力音声を組み立てるために適切な合成単位の情報を設定する設定手段と、音韻情報、音律情報、及び設定手段で設定された情報に基づいて合成単位での接続部におけるターゲットスペクトルを算出する算出手段と、算出手段で算出されたターゲットスペクトルに基づいて記憶手段に格納されている音声合成パラメータから適切な合成素片を選択する選択手段と、選択手段で選択された合成素片を接続する接続手段とを備えているので、大量の音声パラメータを蓄積しておき、音声の合成のために最適な合成素片を抽出して接続することにより出力音声を合成する。その結果、少ない計算量で明瞭性が高くしかも自然性のよい音声を得ることができる。

【図面の簡単な説明】

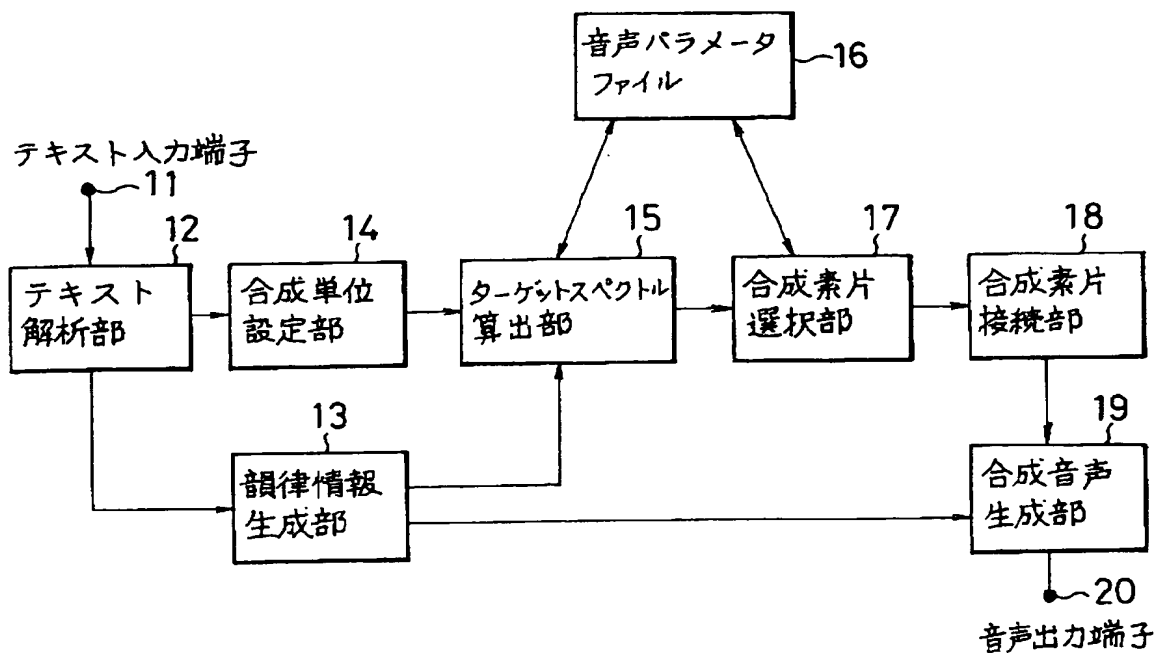
【図1】本発明の音声規則合成装置の一実施例の構成を示すブロック図である。

【図2】図1の音声規則合成装置によるターゲットスペクトルの算出処理を説明するためのフローチャートである。

【符号の説明】

- 11 テキスト入力端子
- 12 テキスト解析部
- 13 韻律情報生成部
- 14 合成単位設定部
- 15 ターゲットスペクトル算出部
- 16 音声パラメータファイル
- 17 合成素片選択部
- 18 合成素片接続部
- 19 合成音声生成部
- 20 音声出力端子

【図1】



BEST AVAILABLE COPY

【図2】

